



# Digital Preservation – Financial Sustainability of Biological Data and Material Resources

Christina Chandras, Thomas Weaver, Michael Zouberakis, John M. Hancock, Paul N. Schofield and Vassilis Aidinis

# INTRODUCTION

- Generation of increasingly diverse and specialized data sets, some extremely large and complex
- As a result some data are put in public databases and bio-repositories
- Databases contain: genomic and metabolomic data, information about structure, function, localization, and clinical effects of mutations and mouse mutants that serve as good models for human disease

***Biological databases have become an important tool in assisting scientists to retrieve information, understand and explain biological molecules and processes***

- Biological knowledge is *distributed worldwide*; it is therefore difficult to:
  - ensure preservation, encourage data deposition and accessibility of information
  - standardize data representation and therefore data quality

# DATA AND BIOLOGICAL RESOURCES

- Scientific progress depends on the ability of researchers to access and exploit data
- However, sharing of data is not always a researchers first priority
- A new concept has been established by US National Academies committee:

***UPSIDE; Uniform Principle for Sharing Integral Data and Materials Expediently***

- *The responsibility of authors to share data and materials referenced in their publications*
- *The role of journals to impose requirements for data and material sharing*

- A central role for BRCs is to champion the principles set out by UPSIDE and embrace the Open Access policy, quality of material, data integration and sustainability

# DATA AND BIOLOGICAL RESOURCES

- Recently published guidelines by the Organization for Economic Co-operation and Development (OECD), highlight that adequate and reliable sources of funding are required to facilitate the sharing infrastructure and long term stability of BRCs
- Evidently, related data needs to be supported, otherwise, valuable information may be lost or transferred into a completely commercial environment
  - Blockade of access to the information and/or high cost
  - Loss of data and potential for technology transfer
- At the same time, as certain data types (e.g. imaging, microarray, phenotypic etc) may include costly processes for their generation, it can be potentially difficult to fulfill scientific duty and make resources available
- In order to achieve this multi-source portal (integration of databases) all contributing BRCs should be:
  - validated for their data/information quality according to accepted standards
  - continuously update both the level of material/data as well as incorporation of novel biological resources
  - support offered by biologists/curators and informaticians

# PROBLEMS ENCOUNTERED

- Although original funding comes relatively easily where the community need is clear
- Commonly databases face loss of original support after a few years in development
- A major problem for most databases – securing financial support for the bioinformaticians and curators who create and maintain them
- Lack of secure funding may frequently result in decommissioning and loss of valuable and irreplaceable data

***Long-term sustainability of databases requires adequate and reliable sources of funding***

***How does one support a useful BRC and can ensure appropriate data/information archival and curation?***

# MODELS EXAMINED

- *Cost Recovery*  
establishment of an annual equitable structure for the standing expenses towards major utilities of the respective biological resource centre (provided by: government programs, non-profit organizations, pharmaceutical firms etc)
- *Fee-for service*  
a standard business model where services are unbundled and paid for separately (pay per view interface – restricted access policy)
- *Institutional Funding*  
allocated funds obtained from a particular institution towards the respective biological resource centre
- *Top Slicing publicly funded grants*  
a percentage of the allocated funds is set aside to ensure long-term accessibility of related data

# THE ROLE OF INDUSTRY VS THE ROLE OF GOVERNMENT

- Both industry and government have proven to comprise the majority of funding bodies supporting BRCs

## ***Government:***

- Government funds provide a successful solution for BRCs, however this is not totally effective since benefits are not always enough to fully cover the infrastructure and business development overheads, leaving BRCs liable to service charges

## ***Industry:***

- Although industry investments are extremely essential since they provide invaluable support towards further development of resource assets, on the other hand the benefits again will not be enough to cover the infrastructure and business development which will outweigh any benefit from the original attempt

***Some BRCs have attained a dual support system***

## A MODEL WITH POTENTIAL: ACADEMIC – COMMERCIAL PARTNERSHIP FOCUS ON CORE COMPETENCIES

- Academic laboratories:
  - are mostly sustained by institutional funding or grants
  - they develop new applications, tools and analysis systems
  - they also support the identification of communal needs and define quality standards
- Commercial organizations on the other hand:
  - financed by their own commercial activities
  - function in a collaborative manner between research and licensing
  - operate as service providers, offering standard technologies and quality systems, sales and marketing distributors

***An examined model that appears to be successful towards prolonged financial sustainability of BRCs is an 'Academic-Commercial partnership'***



# MMdb ([www.mugen-noe.org/database](http://www.mugen-noe.org/database))

- The MUGEN Mouse database (MMdb) serves as a use-case example
- The MUGEN Mouse Database (MMdb) is a virtual online repository of murine models of immune processes and immunological disease
- MMdb is 1 out of 10 virtual repositories **contributing data to IMSR**
- Aidinis *et. al.*, 2008 *Nucl. Acids Res.* **36**: D1048-D1054 (Database Issue)

*D1048–D1054 Nucleic Acids Research, 2008, Vol. 36, Database issue*  
doi:10.1093/nar/gkm838

*Published online 11 October 2007*

## **MUGEN mouse database; Animal models of human immunological diseases**

**V. Aidinis\***, C. Chandras, M. Manoloukos, A. Thanassopoulou, K. Kranidioti, M. Armaka, E. Douni, D. L. Kontoyiannis, M. Zouberakis, G. Kollias<sup>†</sup> and the Mugen NoE consortium<sup>‡</sup>

B.S.R.C. Alexander Fleming, 34 Fleming Street, 16672, Vari, Greece

Received August 10, 2007; Revised September 24, 2007; Accepted September 25, 2007

# MMdb (www.mugen-noe.org/database)

The screenshot displays the MUGEN Mouse Database interface. At the top, there is a navigation bar with links for home, help, contact, about, and search. Below this, a sidebar on the left contains a menu with 'MUGEN Mice', 'Genes', and 'Resources'. The main content area features a 'Welcome' message and a search results table for 'mugen mice'. The table has columns for MUGEN ID, Common Line Name, Mutation(s), Research Application, and Updated date. A search filter is visible above the table, and a login form is located in the bottom left corner. The table lists various mouse models, including M145005 (Mx-Cre), M147002 (LysM-Cre<sup>K/I</sup>), and M141004 (IFN $\gamma$ -Receptor).

**Navigation:** home | help | contact | about | search: mutants | genes | alleles | mp terms | v.2.1.0

**Search Results:** mugen mice

**Filters:** back reset help & info


Research Applications	Gene	Institution	Researcher	Mutation Type	MP Terms	Items Page
*	*	*	*	*	*	40


displaying 1-40 out of 141

MUGEN ID	Common Line Name	Mutation(s)	Research Application	Updated
M145005	Mx-Cre	transgenic	TRANSGENIC TOOL	2007-08-06
M147002	LysM-Cre <sup>K/I</sup>	insertion, targeted mutation	TRANSGENIC TOOL	2007-08-04
M141004	IFN $\gamma$ -Receptor	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M145007	MnSOD	targeted mutation	MODEL OF HUMAN DISEASE	2007-08-06
M141006	pentraxin 3, TSG-14	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M167001	CHORDIN COIN	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-29
M163001	CD2-BTLA	transgenic	MODEL OF IMMUNE PROCESSES	2007-08-28
M193018	MMP-9	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M191001	LDLR	targeted mutation	MODEL OF HUMAN DISEASE	2007-08-07
M193033	Rosa26-Cre-Ert <sup>2</sup>	targeted mutation	TRANSGENIC TOOL	2007-09-26
M193034	Rosa26-Cre-Ert <sup>2</sup>	targeted mutation	TRANSGENIC TOOL	2007-09-26
M189001	P14	transgenic	TRANSGENIC TOOL	2007-09-26
M145002	VI10	insertion, targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-06
M193014	LAT Y7/8/9F	targeted mutation	MODEL OF HUMAN DISEASE	2007-09-26
M211009	Langerin-DTR/EGFP	targeted mutation	TRANSGENIC TOOL	2007-09-26
M139008	IL-1 $\alpha$ K/O (Roer)	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M193037	Pten	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-07
M201009	GFAP-Cre	transgenic	TRANSGENIC TOOL	2007-08-28
M193039	CCL10-RTA	transgenic	TRANSGENIC TOOL	2007-09-27
M211010	SLIT1	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-23
M141002	IL-1 $\alpha$ K/O (Can)	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M225000	DNAH9-DTA	targeted mutation	TRANSGENIC TOOL	2007-09-18
M207002	SIC1	targeted mutation	MODEL OF HUMAN DISEASE	2007-09-26
M195005	CD2-iCre	transgenic	TRANSGENIC TOOL	2007-09-26
M201008	Inducible Beta 1 Integrin KO	targeted mutation	MODEL OF HUMAN DISEASE	2007-08-30
M193005	LAT Y6/7/8/9F	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M193008	NTAL K/O	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M193004	LAT K/O B6	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M193011	SUPER-ZAP	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-04
M193012	CCR9	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M193013	CCL25	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M201001	LIME K/O	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M215012	Foxp3 EGFP	targeted mutation	TRANSGENIC TOOL	2007-08-03
M149001	Dg- $\alpha$	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-09-26
M153001	NALP3	insertion, targeted mutation	MODEL OF HUMAN DISEASE	2007-10-08
M145006	Gp130	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M145008	Beta 7 Integrin	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03
M139001	CD11c-Cre-Ert <sup>2</sup>	transgenic	TRANSGENIC TOOL	2007-08-28
M199001	CD3-e K/O	targeted mutation	MODEL OF IMMUNE PROCESSES	2007-08-03

login form: user: [input] pass: [input] login

# MMdb & DATABASE SUSTAINABILITY





home | help | contact | about | search:
mutants | genes | alleles | mp terms
v.2.1.0

**MUGEN Mice**

[View All](#)

[Genes](#)

[Resources](#)

**IL-10<sup>K/O</sup> (Cgn)**

[back](#) [help & info](#) [expand](#) [collapse](#) [pdf version](#)

---

**General Information**

<b>MUGEN ID</b> M141002	<b>Common Line name</b> IL-10 <sup>K/O</sup> (Cgn)	<b>Research applications type</b> MODEL OF IMMUNE PROCESSES
<b>Contact</b> Muller Werner	<b>Institution</b> University of Manchester	<b>Research applications comments</b> Chronic colitis model

---

**Availability**

Repository	Available Genetic Background	Strain State	Strain Type
<a href="#">JACKSONS</a>	C57BL/6J	live mouse	mutant stock
<a href="#">GBF</a>	C57BL/6J	live mouse	mutant stock

---

**Genetic Background**

DNA Origin	Targeted Background	Host Background	Backcrossing Strain	Backcrosses
neo: bacterial	129P2/OlaHsd	C57BL/6	C57BL/6	>=10

---

**Strain Information**

Designation	JAX ID
B6.129P2-Il10 <sup>tm1Cgn</sup>	002251

---

**Allele & Mutations**

Symbol	Name	Mutation Types
Il10 <sup>tm1Cgn</sup>	targeted mutation 1, University of Cologne	targeted mutation

---

**Genes Affected**

Symbol	Name	Chromosome
Il10	interleukin 10	1

---

**Handling & Genotyping Instructions**

Name	Type
<a href="#">Links</a>	
<a href="#">Genotyping Protocol</a>	Weblink

---

**Mammalian Phenotype Ontologies**

**MP Term:** [digestive/alimentary phenotype](#)

**PATO Logical Definition:** [quality](#) inheres\_in [digestive system](#)

---

**MP Term:** [digestive/alimentary phenotype](#)>[abnormal digestive system physiology](#)>[intestinal inflammation](#)>[large intestinal inflammation](#)>[colitis](#)

**PATO Logical Definition:** [increased](#) inheres\_in [inflammatory response](#) has\_central\_participant [colon](#)

---

**References**

Name	Type	Comment
<a href="#">Kuhn et al., 1993 Cell 75(2):263-74</a>	Webblink	Interleukin-10-deficient mice develop chronic enterocolitis.

---

**Author's Comments**


These IL-10-deficient mice, generated by gene targeting, have normal lymphocyte development and antibody responses, but most animals are growth retarded and anemic and suffer from chronic enterocolitis.










---

**User's Comments** [\[post comment\]](#)

---

[back](#)



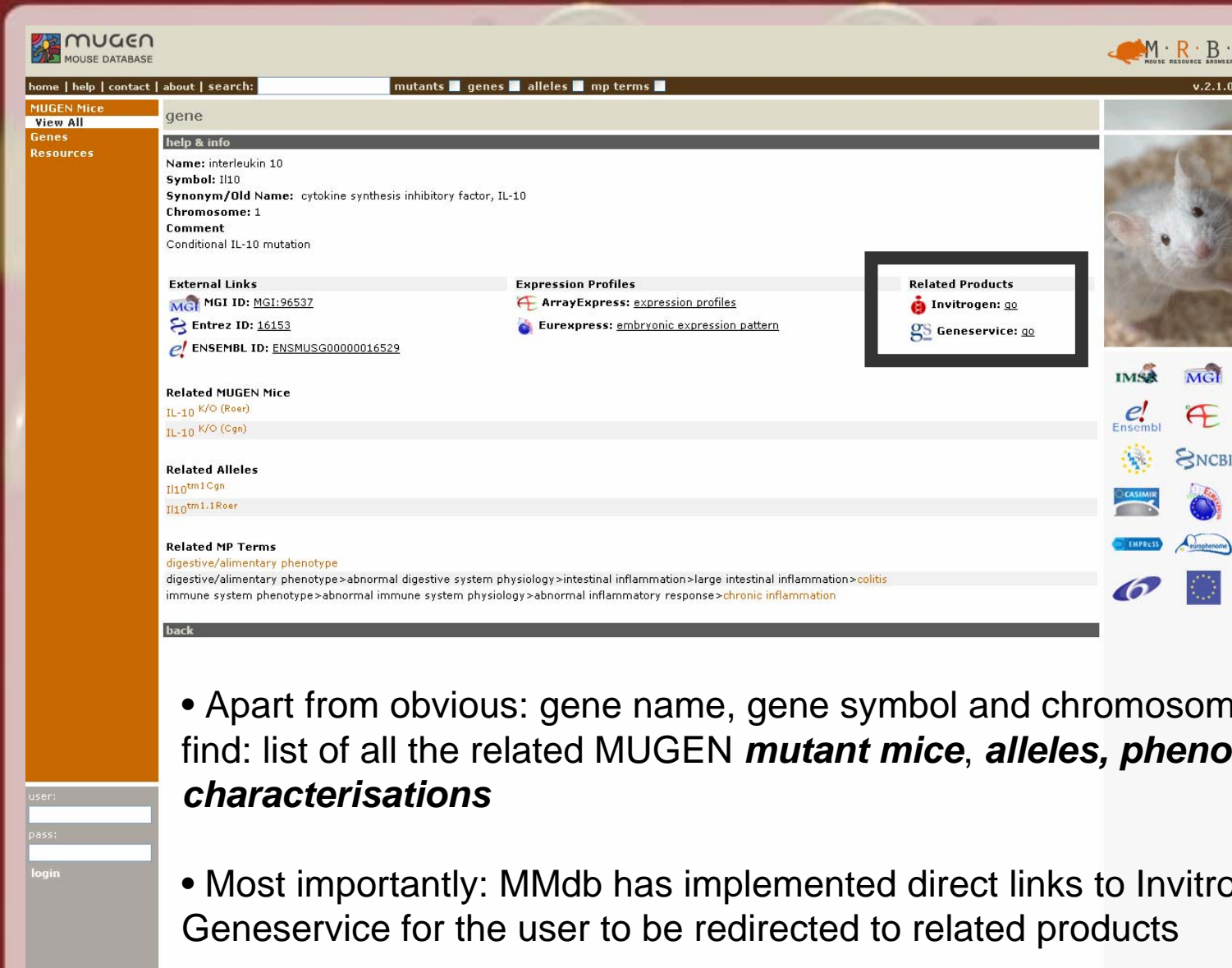










user:

pass:

[login](#)

# MMdb (www.mugen-noe.org/database)



The screenshot displays the MUGEN Mouse Database interface for the gene 'interleukin 10'. The page is structured with a navigation bar at the top, a search bar, and a sidebar on the left. The main content area is divided into several sections: 'help & info', 'External Links', 'Expression Profiles', 'Related Products', 'Related MUGEN Mice', 'Related Alleles', and 'Related MP Terms'. A red box highlights the 'Related Products' section, which contains links to 'Invitrogen: go' and 'Geneservice: go'. The 'Related MUGEN Mice' section lists 'IL-10 K/O (Roer)' and 'IL-10 K/O (Cgn)'. The 'Related Alleles' section lists 'Il10<sup>tm1Cgn</sup>' and 'Il10<sup>tm1.1Roer</sup>'. The 'Related MP Terms' section lists 'digestive/alimentary phenotype', 'digestive/alimentary phenotype>abnormal digestive system physiology>intestinal inflammation>large intestinal inflammation>colitis', and 'immune system phenotype>abnormal immune system physiology>abnormal inflammatory response>chronic inflammation'. A 'back' button is located at the bottom of the main content area. The sidebar on the left contains a 'MUGEN Mice' section with a 'View All' link, and a 'login' section with 'user:' and 'pass:' input fields and a 'login' button. The top right corner of the page shows the 'M·R·B·' logo and the version number 'v.2.1.0'.

- Apart from obvious: gene name, gene symbol and chromosome, one can find: list of all the related MUGEN **mutant mice, alleles, phenotypic characterisations**
- Most importantly: MMdb has implemented direct links to Invitrogen and Geneservice for the user to be redirected to related products

# MMdb (www.mugen-noe.org/database)

The screenshot shows the Invitrogen LINNEA ONLINE GUIDES interface. The search results for 'BTD' are categorized into several sections:

- Antibody Target:** A table listing various antibody products for BTD, including details like Catalog #, Size, and List Price.
- Assay:** A table listing various assay kits for BTD, including details like Catalog #, Size, and List Price.
- Labeling and Detection:** A table listing various labeling and detection reagents for BTD.
- PCR Product:** A table listing various PCR products for BTD.
- Protein:** A table listing various protein products for BTD.
- RNAi:** A table listing various RNAi products for BTD.

Antibody Target

Assay

Labeling and Detection

PCR Product

Protein

RNAi

- MMdb has approached Invitrogen and other potential companies, asking them to:
  - link their individual products with the respective mouse model
  - examine the possibility that they would be interested in linking with MMdb
  - explore their willingness towards marketing/advertisement service charges which could help maintain the databases

• Although Invitrogen responded positively, the overall response was not very successful

• This approach although appearing with great potential in theory, in practice is harder to achieve as companies are not that willing to sponsor academic institutions

• This may be a matter of time and should companies be appropriately primed this arrangement may indeed prove to be beneficial towards prolonged financial sustainability

# THE ROLE OF CONSORTIA

- The EU in support of FP5 and FP6 sponsored a number of projects generating biological experimental data
- Some of these consortia also serve as liaisons towards the European Commission, giving advice with respect to specific areas of interest and their respective needs for further development, suggesting potential future directions that the EU should pursue
- The European Commission has also supported some CA (e.g. PRIME & CASIMIR) especially to organize and bring together the individual European efforts as well as survey the scientific community needs
- This interactive relationship allows networks to lobby both national and international funders and potentially improve application practices and for funders to approach and consult with the network with regard to issues and priorities

***Consortia play an intermediary role between the scientific community and the European Commission***

# CONCLUSION

- Data must be securely stored and freely available to the research community
- BRCs should not exist as data warehouses, but rather a cluster of activities supporting the community of academic and commercial researchers
- Long term sustainability of BRCs requires adequate and reliable sources of funding so that data is preserved and disseminated properly
- With regard to the examined business models:
  - the “full cost recovery” and “fee-for-service” models have some flaws
  - most promising are the “institutional funding” and “top slicing of public funding” models both of which provide a secure environment for the BRC
  - the academic-commercial partnership may appear to have potential should vendor corporations become involved in this collaborative effort

***Prolonged financial sustainability is vital for data preservation and development of cross-querying***

# ACKNOWLEDGMENTS

*Thank you for your attention*

## *MUGEN & CASIMIR*

Vassilis Aidinis

Michael Zouberakis

Paul Schofield

John Hancock

Thomas Weaver

## *COLLABORATORS*

MUGEN

CASIMIR

CREATE

I-DCC

